DOCUMENT RESUME

ED 171 728 TM 008 592

AUTHOR Haenn, Joseph F.; Proctor, Diane C.
TITLE A Practitioner's Guide to Out-of-Level Testing.
INSTITUTION National Testing Service, Inc., Durham, N.C.
PUB DATE Mar 78
NOTE 50p.; Paper presented at the Annual Meeting of the American Educational Research Association (62nd, Toronto, Ontario, Canada, March 27-31, 1978); Pages A3-A10 may not produce legibly due to small print size of original document

EDRS PRICE MF01 Plus Postage. PC Not Available from EDRS.
DESCRIPTORS *Achievement Tests; Compensatory Education Programs; Educationally Disadvantaged; Elementary Secondary Education; Equated Scores; *Grade Equivalent Scores; Instructional Program Divisions; *Low Achievers; Student Ability; *Student Testing; *Testing Problems; Test Interpretation; *Test Selection
IDENTIFIERS Elementary Secondary Education Act Title I; *Out of Level Testing; Title I Evaluation and Reporting System

ABSTRACT

Out of level testing is defined and reasons for administering out of level tests, especially in compensatory education programs are explained: (1) to achieve optimum test reliability; (2) to test student achievement--provided out of level test scores are interpreted in terms of the appropriate in-level population; (3) to avoid carelessness and inattentiveness or frustration and guessing caused by students' taking tests at an inappropriate level; and (4) to avoid large numbers of students scoring at chance level. Two procedures which allow out of level tests to be converted to in-level scores are described using publisher produced raw score-to-percentile conversion tables, or standard score-to-percentile conversion tables. Chance level scores, names of expanded standard score scales, and worksheets to convert out of level scores are appended. (MH)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

NATIONAL TESTING SERVICE, INC.
1905 Chapel Hill Road
Durham, North Carolina 27707

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

ED171728

A PRACTITIONER'S GUIDE TO OUT-OF-LEVEL TESTING

Joseph F. Haenn
Diane C. Proctor

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

Joseph F. Haenn

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC) AND USERS OF THE ERIC SYSTEM."

NTS Research Corporation
1905 Chapel Hill Road
Durham, N.C. 27707

A paper prepared for presentation at the Annual Meeting of the American Educational Research Association in Toronto, Canada.

March 1978

The opinions expressed in this paper are strictly those of the authors and do not reflect any policy, official or unofficial, of NTS Research Corporation or any other organization.

TM008 592

A current trend in educational measurement and evaluation is to test students at their current level of instructional functioning rather than at their actual grade placement. For example, a fifth grade student reading at the second grade level would have a difficult time completing a test designed for fifth grade readers. Test publishers generally develop several contiguous levels (e.g., second grade level, third grade level, etc.) or overlapping levels (e.g., second and third grade level, third and fourth grade level, etc.) of tests nominally intended for a particular grade (or age) or range of grades (or ages). Administering the test level in the grade (age) nominally recommended by the publisher is referred to as in-level or grade (age) level testing. Out-of-level testing refers to the practice of administering the test level at a grade (age) other than that nominally recommended by the publisher. Following the example above, administration of a test nominally recommended only for the second grade to the fifth grade student would be called testing out of level. This procedure is also sometimes referred to as instructional level or functional level testing, especially when the test level is selected for classrooms or groups of students as a whole. If test levels are chosen appropriate to each student, this process is sometimes also known as individualized testing.

Out-of-level testing generally refers to testing either above or below the nominally recommended level. The authors' special concern is with testing compensatory education students, especially in light of the recent introduction of the Title I Evaluation and Reporting System proposed by the U.S. Office of Education. Since most compensatory education students are functioning below the expectations for their grade (age) groups, this paper will concentrate on testing below the level nominally recommended by test publishers.

The concept of out-of-level or functional level testing has been around for over fifteen years. During this time the concept has often been misunderstood and even misused. In fact, two states currently prohibit testing students other than in level. However, the practice has received new impetus with the emergence of the proposed Title I Evaluation and Reporting System. This system recommends testing students at their functional level, which quite often implies out-of-level (below grade level) testing rather than nominal grade level, or in-level, testing.

The following section of this paper reviews the significant literature and events to date which have had impact on the concept of out-of-level testing. Subsequent sections explore the reasons for testing out-of-level, the procedures which need to be followed in conducting out-of-level testing and the steps which must be taken to correctly convert obtained out-of-level scores to the appropriate in-level norms. Three appendices supplement the text. Appendix A presents chance level scores and their percentile and grade equivalents for the more commonly utilized achievement tests. Appendix B provides the names of the expanded standard score scales utilized by these tests and Appendix C presents two alternative worksheets which can be used to convert to in-level scores.

Review of the Literature

A survey of the current literature revealed a very limited number of studies dealing with out-of-level testing. The earliest study was a 1961 dissertation by Fisher, who used out-of-grade tests with retarded and accelerated readers. He found out-of-level tests to be better measures of these students' abilities and to yield approximately the same mean grade equivalent score as the in-level test using the Iowa Tests of Basic Skills (ITBS). Fisher's study was cited as providing a basis for the move to out-of-level testing in 1968 by the city of Philadelphia (Ayrer and MacNamara, 1973).

Philadelphia (Ayrer and MacNamara, 1973), Chicago (Wick, Note 1 and Note 2), Dallas (Mendro, Note 3), and Atlanta (Barnes, Note 4) have all utilized the Iowa Tests of Basic Skills (ITBS) in their district-wide testing programs and have provided for out-of-level testing. The findings in each of these cities has been remarkably similar. The percentage of students scoring at or below the chance or guessing level (defined as the number of test items divided by the number of options per item) decreased substantially when students were tested out-of-level. However, the mean grade equivalent scores were also lower. In general, these school districts seem to have accepted the lower mean scores in exchange for scores regarded as being more valid and reliable. However, these results do raise a question as to the comparability of in-level and out-of-level test scores. At a 1977 Houghton Mifflin conference focusing on out-of-level testing, Hieronymus addressed this issue in stating: "We expect people on the average to get the same derived scores from taking different levels...The purpose of [out-of-level] testing

is not to obtain systematically higher or lower scores, but a more accurate, valid and effective measurement" (Hieronymus, Note 5, p. 25).

In Chicago (Wick, Note 2), out-of-level testing is resulting in substantial year-to-year gains being evident in test scores, most notably with their lowest performing students. Wick states: "...had we not gone to individualized testing and gotten rid of those chance scores and noise, we would not have seen that gain. It would have been so hidden in the noise and the floor level of the test that we couldn't have picked it up...." (p. 38).

A major shortcoming with the aforementioned data reports (Ayrer and MacNamara, 1973; Fisher, 1961; Wick, Note 1 and Note 2; Mendro, Note 3; Barnes, Note 4) is that none employed a research design which allowed direct comparison of in-level and out-of-level scores for the same point in time. (Fisher's data were based on in-level data collected in January and out-of-level data collected in April). Long, Schaffran and Kellogg (1977), however, did carry out a study in which in-level and out-of-level testing could be compared. Long et al. (1977) administered both the nominally recommended level and a level corresponding to the child's instructional reading level to a sample of second, third, and fourth grade Title I students in Rhode Island. At the second and third grades, they found that tests administered one or two levels below the nominally recommended level resulted in higher mean grade equivalent scores than tests administered in-level. At the fourth grade, the reverse trend was observed. Young (1972) reported similar results for the Total Reading subtest of the Metropolitan Achievement Tests (MAT). At the third grade no differences were found. At the fifth and seventh grades, the in-level test

resulted in higher mean grade equivalent scores than the out-of-level test. Pelavin and Barker (Note 6) have also reported a lack of level-to-level score comparability on the MAT.

Long et al. (1977) also looked at their data from the standpoint of its effect on the selection of students for Title I programs and its effect on impact evaluation. They found that a fairly large proportion (9% at grade four to 43% at grade two) of students met the eligibility criteria based on one test level but not the other. When pre-to-posttest changes were examined, it was found that, with the exception of the third grade vocabulary subtest, the magnitude of change was greater with instructional level tests than with grade level tests.

Out-of-level testing has also been used with educable mentally retarded (Yoshida, 1976) and with academically gifted students (Keating, 1975). In both of these studies the issue was not score comparability, but rather selecting test levels at which a child's performance could be validly and reliably assessed.

The developer of the Title I Evaluation and Reporting System, RMC Research Corporation, has been among the primary contributors to the out-of-level testing literature. Three of their technical papers (Roberts, 1976; Tallmadge and Roberts, Note 7; and Roberts, Note 8) address issues related to out-of-level testing, particularly floor and ceiling effects and their impact on program evaluations. These papers are primarily directed toward out-of-level testing in compensatory education programs.

Until the publication of the proposed Title I Evaluation and Reporting System, few test publishers addressed the issue of out-of-level or functional level testing. Although most of the widely used achievement tests provide expanded standard score scales which enable one to test out-of-level, few

test manuals discuss out-of-level testing. Among these tests, out-of-level testing is mentioned only for the Iowa Tests of Basic Skills (ITBS), the Metropolitan Achievement Tests (MAT), the 1977 edition of the California Achievement Tests (CAT), forms C and D, and a set of special out-of-level norms which were prepared for the Gates-MacGinitie Reading Tests in 1977. These test manuals endorse out-of-level testing, but provide few firm guidelines for selecting appropriate test levels and provide virtually no instructions for determining appropriate in-level norm referenced scores for tests administered out-of-level.

Of the major achievement batteries, the CAT was the first to publish new forms since the introduction of the proposed Title I Evaluation and Reporting System. Whether the impetus was the proposed evaluation requirements for Title I or a general trend in American education, CAT forms C and D (1977) have placed considerable emphasis on out-of-level or functional level testing.

Noting that publishers like to have products which are marketable, and noting that Title I is a potentially large market place, it seems likely that other publishers will be following suit and addressing out-of-level testing in their test manuals. Hopefully, in the course of this movement toward out-of-level testing, research will be done so that we know more about such issues as score comparability. Also, it is hoped that test publishers will provide information to assist the test user in choosing appropriate test levels and in correctly interpreting scores from out-of-level tests.

Reasons for Testing Out-of-Level

The optimum test reliability is obtained when students correctly respond to slightly more than half of the test items (Stanley and Hopkins, 1972). Under these conditions there is no reason for administration of an out-of-level test as the in-level test is maximally measuring student achievement on this set of items. However, when a significantly large number of students within a testing group correctly respond to considerably more or fewer items, the reliability of the test for this group of students diminishes. The relationship between reliability and percentage of correct items on a four-choice multiple choice test is roughly illustrated in Figure 1.

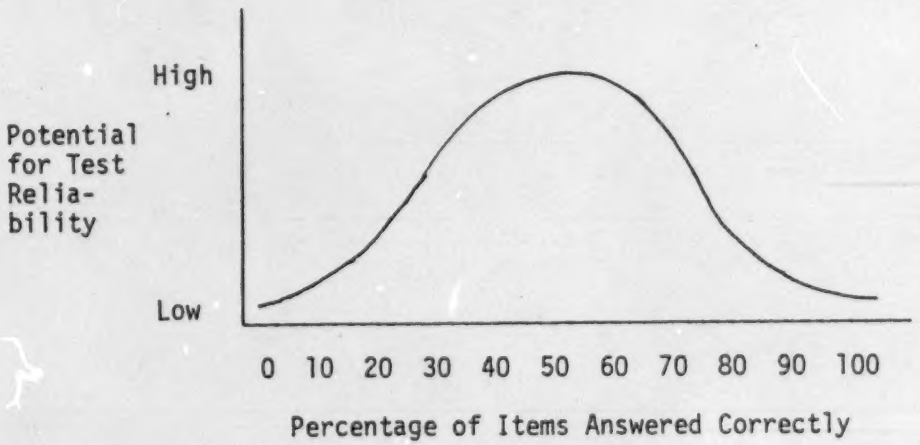


Figure 1. Relationship between percentage of correct items and test reliability.

A common misconception regarding out-of-level testing is that it gives students who are functioning below grade level an unfair advantage since they are administered an easier level of the test and will earn higher scores. However, this belief is invalid if the out-of-level score is interpreted in terms of the appropriate in-level reference or normative population.

The major problem with out-of-level testing to date has been the utilization of an inappropriate reference population, i.e. the out-of-level norm group rather than the in-level norm group. A student taking a lower (easier) level of the test may obtain a higher raw score than for the in-level test. However, the student must answer more items correctly on the out-of-level test than on the in-level test to obtain the same norm-referenced score (e.g., grade equivalent, expanded standard score, percentile rank). A later section of this paper presents the proper procedures for the correct scoring conversions of out-of-level raw scores to in-level norm referenced scores.

The remainder of this section of the paper explains some additional conditions under which out-of-level testing is appropriate.

Ceiling Effects

If students find that items are too easy they may become inattentive and make careless mistakes. A general rule of thumb to determine whether a test is too easy for a given group of students is to observe whether the mean of the group exceeds a score equivalent to three-quarters or more of the raw score items correct (Roberts, 1976). If this occurs a "ceiling effect" may have been encountered.

A ceiling effect implies that a large proportion of the students are reaching their maximum possible score on the test. This maximum score may be less than the total number of items due to carelessness or misleading item structure. A convenient check to determine whether a ceiling effect has been encountered for a group of students is to compare the mean of the respondent group with their median score (Roberts, 1976). If the mean is at least one-third of a standard deviation lower than the median, then the

chances are very good that these students have "topped out" on this test level. The responses illustrated in Figure 2 demonstrate a ceiling effect.

If a large proportion of students are "topping out" on the test either at pretest or posttest time, then the potential for demonstrating student gains between pretest and posttest when they actually occur is reduced. This happens because a ceiling effect systematically underestimates the actual level of student performance. For example, if the group responses from Figure 2 are used, a large proportion of the students (15 percent) are indicated as answering all ten items correct. If a longer test had been given with only some more difficult items added, the chances are great that many of these students could have obtained an even higher score. Even those students with nine or fewer correct items could have increased their scores due to a combination of level of knowledge and chance responding. Therefore,

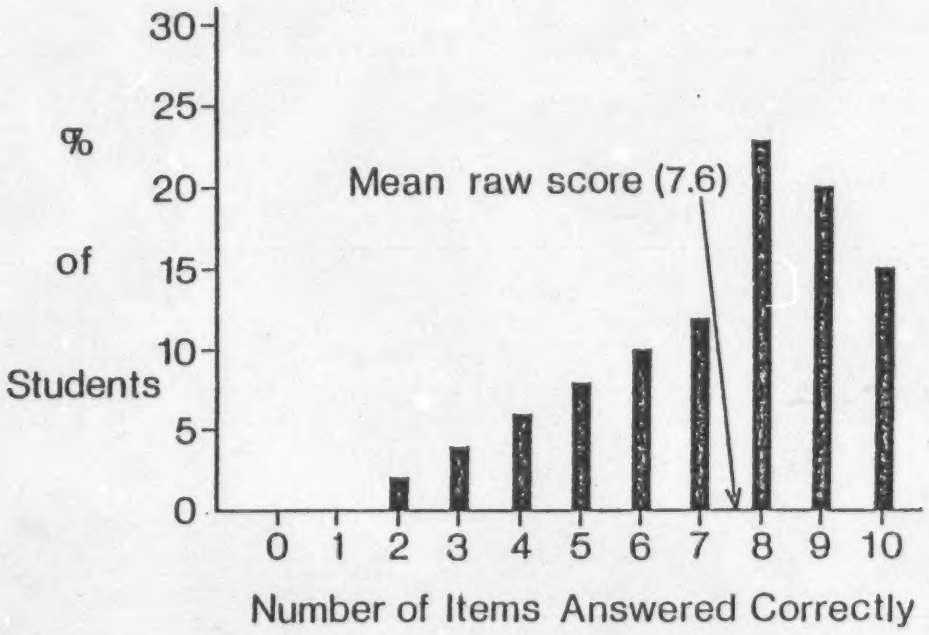


Figure 2. The ceiling effect which results from administering a test which is too easy.

we might be underestimating potential student performance at posttest time (point A in Figure 3) when compared to obtained performance (point B) and reduce the potential for student gains (difference between point A and point B when compared to point C). It should be obvious that a ceiling effect at pretest can be even more serious and result in either no growth or negative growth if the same test level is used again at posttest time.

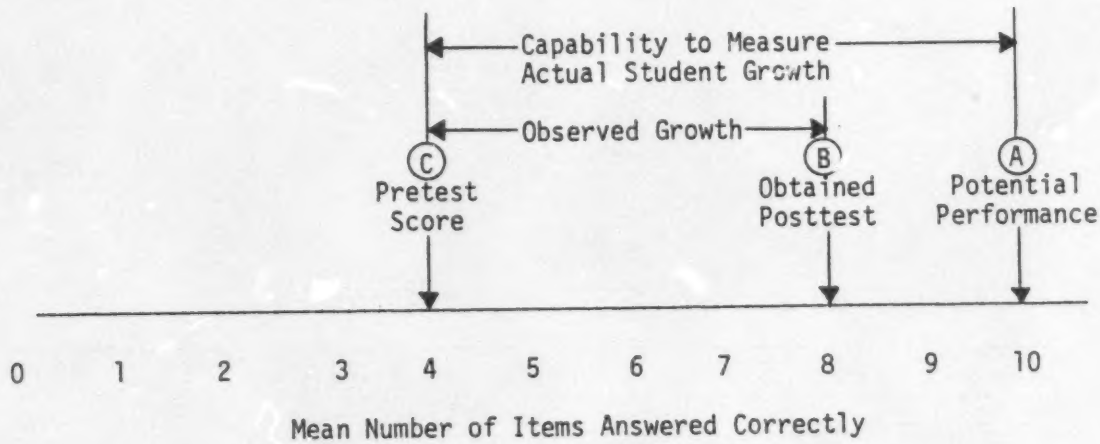


Figure 3. Reduced potential for student gains due to ceiling effect at posttest time.

It is not the purpose of this paper to discuss ceiling effects in any detail. However, they do occur even with educationally disadvantaged student populations. For a further discussion of this topic, see Roberts (Note 8).

Floor Effects

Students who can answer only a few items correctly will tend to become frustrated and, on closed-response items such as true/false or multiple choice items, may respond by guessing--often by marking responses without reading the questions. A general rule of thumb to determine whether a test is too

difficult for a given group of students is to determine whether the mean of the group exceeds a score equivalent to one-third or less of the raw score items correct (Roberts, 1976). If this occurs, then a "floor effect" may have been encountered.

A floor effect implies that a large proportion of the students are scoring at the minimal score possible on the test. This minimal score is usually considerably greater than simply no correct items except on open-ended response items, such as completion or fill-in-the-blank items. On closed-response items the minimal score is usually indicated as the chance score which is as discussed in the next section. A convenient check on the floor effect is to compare the raw score mean of the respondent group with their median score (Roberts, 1976). If the mean is at least one-third of a standard deviation higher than the median, then the chances are very good that these students have "bottomed out" on this level. The responses illustrated in Figure 4 demonstrate a floor effect.

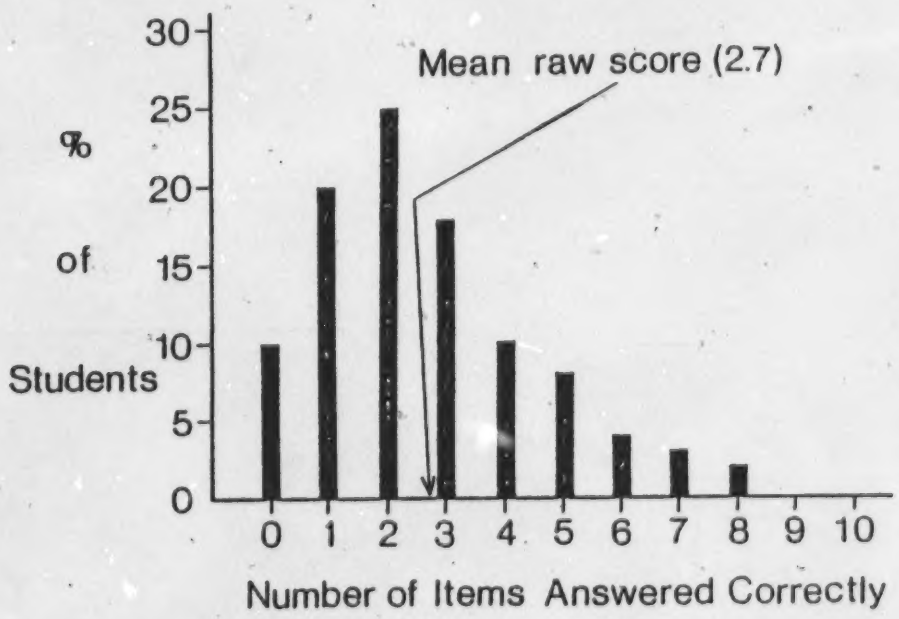


Figure 4. The floor effect which results from administering a test which is too difficult.

If a large proportion of students are "bottoming out" on the test either at pretest or posttest time, then the potential for demonstrating actual student gains between pretest and posttest is reduced. This occurs because a floor effect systematically overestimates the actual level of student performance. For example, if the group responses from Figure 5 are used, a large proportion of the students (55 percent) scored below the chance level score (assuming four-response, closed ended items). If we had given a longer test with only some easier items added, there is a high probability that many of these students would have missed even more items before arriving at their true score. Those students on the shorter test who were not functioning at the chance level purely by chance would then have higher scores on the longer test when compared to those students who were functioning purely at chance level. Therefore, we might be overestimating student potential performance at pretest (point A in Figure 5) when compared to obtained performance (point B) and reduce the potential for student gains (difference between point A and point B when compared to point C). It should be obvious that a floor effect at posttest probably will result in either no growth or negative growth if the same test level is used again as was used at pretest time.

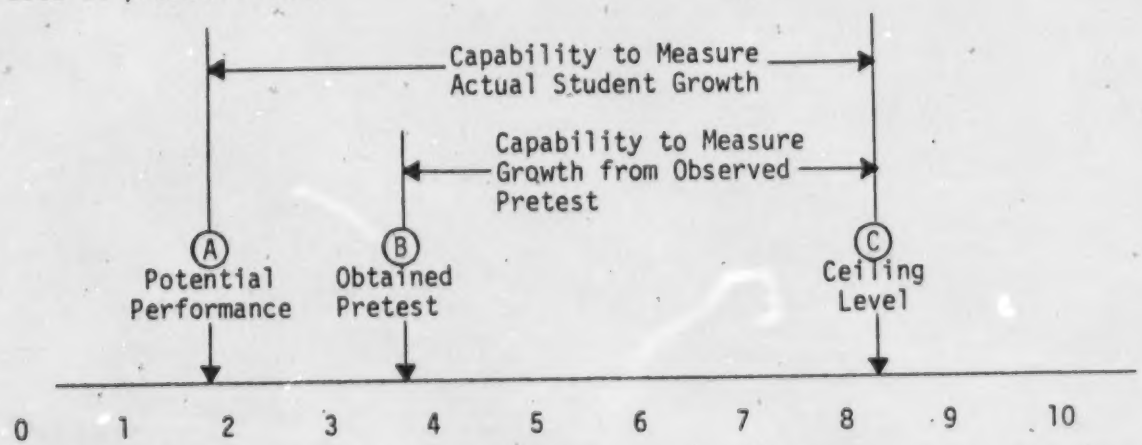


Figure 5. Reduced potential for student gains due to floor effect at pretest time.

Chance Scores

One of the common characteristics of a group of students who encounter a floor effect is a substantial number of students scoring at or around the chance level. The pure chance level of a test is determined by dividing the number of items (N) by the number of alternatives per item (A) (Gulliksen, 1950) (See Note 9). Mathematically, this is expressed as:

$$\text{Test Chance Score} = \frac{N}{A}$$

For example, on a 60 item four-choice, multiple choice test the chance score would be 15 raw score points (60÷4). On a 30 item true/false test the chance score would also be 15 raw score points (30÷2). For tests having items with different numbers of alternatives, compute the chance score for each type of item and add these part chance scores together.

On tests where a large number of students score at the chance level, there is very little useful information obtained. The general rule for determining the chance score for the mean of a group of students is also derived from the number of alternatives per item. It is obtained by dividing the number of items (N) by one less than the number of alternatives per item (A-1) (Roberts, Note 8). Mathematically, this is expressed as:

$$\text{General Rule for determining Chance Score for a group mean} = \frac{N}{A-1}$$

For four-choice items, this reduces to the previously given rule of thumb of one-third of the maximum raw score. For five-choice items, the rule of thumb would be one-quarter of the maximum raw score. However, this rule of thumb breaks down for three-choice items because it predicts an unrealistically high chance mean score (the chance score would be

one-half of the maximum raw score) and especially for two-choice items (the chance score would be one hundred percent of the maximum raw score, or a perfect score).

The chance scores on many of the available subtests in reading and mathematics are given in Appendix A for tests which are commonly used with children in grades 2-6. These tables should be examined in some detail when considering use of a given level for a particular group of educationally disadvantaged students. For example, for those tests with Total Reading scores, the chance score at the lowest designated level for in-level administration between grades 2 and 6 ranges from the second to the 39th percentile, with a median percentile of 10. For those tests with Total Mathematics scores, the chance score at the lowest designated level for in-level administration between grades 2 and 6 ranges from the first to the 39th percentile, with a median percentiale of 6. Therefore, even though students may be functioning at the chance level they may still be as high as the 39th percentile according to the tabled norms.

Because of this phenomena, special attention should be directed towards what can happen if test levels are changed from pretest to posttest. For example, on one test if the level which is nominally recommended for the beginning of the second grade is administered at pretest and the next higher level, which is nominally recommended for the end of the second grade, is administered at the posttest the student average could jump from the 21st percentile to the 35th percentile on the Reading Comprehension subtest even though the students would be scoring at the raw score chance level on each test level. On another test by administering the level of the Total Mathematics subtest which is nominally recommended for the beginning of grade 4 as the pretest and the next

higher level which is nominally recommended for the end of grade four as the posttest, the student average at the raw score chance level would rise from the first percentile to the seventh percentile. Therefore the raw score chance levels and the corresponding grade equivalents and percentiles should be carefully examined for the grade(s) of interest (See Note 10).

Procedures for Testing Out-of-Level

Any test can be administered out-of-level if given to a population other than that for which it is nominally recommended. However, in order to appropriately interpret scores from administration of an out-of-level test a procedure must exist whereby the obtained out-of-level raw score can be compared with in-level norms. For example, a hypothetical reading test is designed so that third grade students take Level III, fourth grade students take Level IV, and fifth grade students take Level V. These are the nominally recommended, or in-level, test levels. If a particular group of fifth graders was reading near the third grade level, the nominally recommended level (Level V) would probably be too difficult for these students and might result in chance level scores for a large proportion of the students. In order to avoid this possibility we might test these fifth grade students at their functional level by administering Level III. The obtained raw score would be an out-of-level raw score which would not be meaningful when compared with in-level raw scores obtained by fifth grade students who took the nominally recommended level (Level V). It would also be meaningless to compare this out-of-level raw score to raw and normative scores for third grade students since their raw and normative scores also were

obtained through administration of an in-level (Level III) test. Therefore, a procedure must exist which allows an out-of-level raw score to be interpreted in-level (i.e., a method of taking the Level III out-of-level raw score in-level to Level V so these fifth graders can be compared with other fifth graders).

There are two procedures available which allow out-of-level raw scores to be converted to in-level scores. The first type consists of tables of norms which convert directly from an out-of-level raw score to an in-level score. These conversion tables are generally available for only one level out-of-level, most often only one level below. Using our hypothetical reading test for grade four, the publisher would provide the in-level norms table for Level IV plus out-of-level norms tables for Level III (one level below) and possibly Level V (one level above). These norms tables have the convenience of requiring a single table lookup to convert an out-of-level raw score to an in-level score, but restrict the range of out-of-level testing in many instances.

The second type of conversion procedure consists of the development of an expanded standard score scale. This scale is designed by the publisher to bridge all levels of the test. For example, the publisher of our hypothetical reading test may design an expanded standard score scale to range from 200 to 800. An expanded standard score of 500 would be equal to different percentile equivalents for the different test levels as illustrated in Figure 6.

The exact procedures for converting out-of-level raw scores to in-level scores will be described in the next section. The remainder of this section will present potential methods for the administration of out-of-level tests.

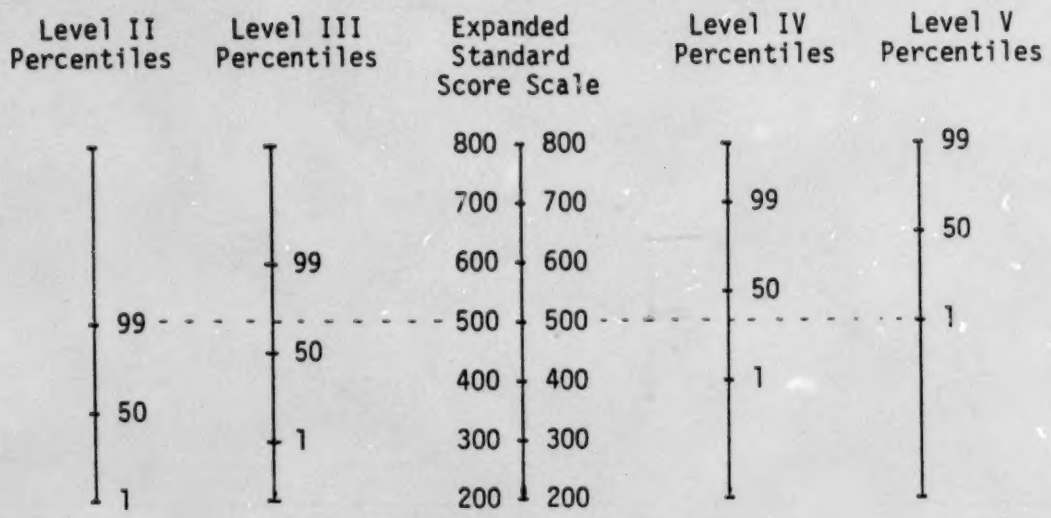


Figure 6. The concept of an expanded standard score scale illustrating the cross-test level equivalencies.

Administration of Out-of-Level Tests

There are basically two methods of testing students at their functional level: single level testing and multiple level testing, with two variations of the latter method. The methods for selecting students and requirements for conducting each type of out-of-level testing are presented below.

Selecting Test Levels Appropriate for Students. There are several sources of information which can be used to select the appropriate test level for a group of students. If scores from the previous year or a previous test administration are available they should be examined carefully. If the mean of the scores for the student sample of interest is less than one-third of the total raw score possible, then the previous test was probably too difficult and the current test should be selected accordingly. If the mean of the previous test scores is three-quarters or more of the total raw score possible, then the previous test was probably too easy for this group of students and the current test should be selected accordingly.

A second source of information should be teacher judgment. Teachers or other specialists should be consulted and should review the content and difficulty of the various test levels to determine the appropriate test level for a given group of students. The level(s) selected should be relevant to student needs, abilities and stage of development. This is especially important in mathematics where grade placement tends to be more rigid.

Some test publishers provide short locator tests to aid the examiner in determining the appropriate level for students. If over three-quarters of the students are designated to take the same level, then that test level may be appropriate for the entire group as the number of students encountering floor or ceiling effects will be minimized. The final determination of whether one level is appropriate for a given group must be a local decision.

Single Level Testing. In many instances an entire group of students (e.g., a group of Title I students) is very homogeneous but considerably different from their grade level peers. In these cases it may be appropriate to administer the same out-of-level test to the entire group of students. This out-of-level procedure is the easiest to administer with the fewest restrictions. While a few students may still be taking a test level which is too difficult or too easy for them, a level should be chosen so that, if possible, at least three-quarters of the students are scoring above the chance level but below the test ceiling.

A word of caution about single level testing is in order. If the test results are to be used for programmatic evaluation or as an initial indicator of individual student achievement, then all students need not be tested at

their functional level. In other words, under these conditions it is not necessary to attempt to have all students scoring above the chance level but below the test ceiling. This is especially so if only one subtest (or a small group of subtests such as those comprising the total reading score) is administered. Most commonly utilized general achievement tests are designed to be highly reliable at the group level but less so at the individual student level. Therefore, it is recommended that a staging concept of testing be employed whereby an achievement test is administered to obtain a group mean to be used for program evaluation and that individual scores be considered only as a rough approximation of individual student status. In a subsequent testing a more specialized diagnostic instrument could be administered to pinpoint individual strengths and weaknesses. In this way a few students scoring near the chance level or test ceiling would not greatly affect the programmatic evaluation, but the chance score information could be utilized to identify those students most in need of further diagnosis (See Note 11).

Multiple Level Testing. Although single level out-of-level testing is the easiest to administer, it often may be impractical. For example, with a regular classroom containing average, accelerated and disadvantaged students, at least three levels of a test may have to be administered: in-level, out-of-level one level above and out-of-level one level below. Even within a group of disadvantaged students more than one level of a test may be appropriate depending upon the heterogeneity of ability of the students and the grade range and content coverage designated for each test level. With most tests the easiest way to accomplish administration of multiple test levels is by dividing the students into a number of groups, with one group for each test level. Then all students who will be taking the same level can be tested at the same time.

A procedure called Individualized Testing is a special instance of multiple level testing. At least two publishers provide multiple test levels with identical instructions, practice items and timing considerations for each level, even though the number of items may differ from one level to the next. In this way multiple levels assigned on an individual basis can be administered at the same time to a heterogeneous group of students.

Steps to Convert to In-Level Scores

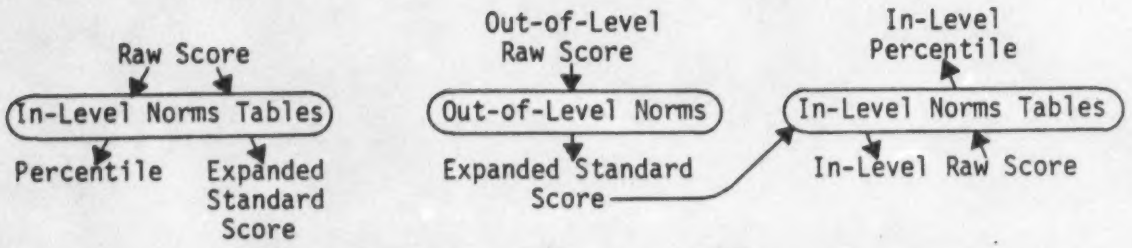
Most of the scoring services provided by test publishers correctly report in-level percentiles for out-of-level test administrations provided the answer documents have been packaged correctly with the appropriate cover sheet. However, many districts hand score their own answer sheets, especially for program evaluations such as Title I. Learning the correct procedures for converting out-of-level raw scores to in-level percentile ranks has been a problem in some school districts.

As noted earlier it is incorrect to use an out-of-level raw score or the norms tables for the out-of-level test when interpreting results obtained out-of-level. Using the hypothetical reading test again, if fifth-graders took the third grade test (Level III) instead of the nominally recommended fifth grade test (Level V), the resultant score would be an out-of-level raw score. This out-of-level raw score must then be converted to the in-level norms.

If the publisher provides out-of-level tables of the type discussed on page 16, the conversion is a simple table lookup of the in-level percentile which corresponds to the out-of-level raw score. However, in most instances publishers instead provide an expanded standard score scale as described

on pages 16-17. There are two variations on the procedures for converting to in-level percentiles. In both instances, the first step is to convert the out-of-level raw score to the expanded standard score provided by the publisher. (See Appendix B for a list of the names of these expanded standard scores for the most commonly utilized tests.) These expanded standard scores cover all of the test levels and theoretically allow the raw scores on any level to be converted to their corresponding raw scores on another level. In this way any test can be administered out-of-level with the resultant scores converted to in-level norms.

Variation I. This variation is used when publishers provide raw score-to-percentile conversion tables. In this variation when testing in-level the raw score can be directly converted to its percentile equivalent. (Figure 7a). However, when out-of-level testing is conducted the out-of-level raw score should be converted to an expanded standard score using the out-of-level (actual administration level) norms. Once the expanded standard score has been determined, it is entered into the in-level (nominally recommended level) norms, so that the in-level raw score corresponding to the out-of-level raw score can be determined. Theoretically, the in-level raw score is the raw score which would have been obtained had the in-level test been administered and no floor or ceiling effects encountered. Once the in-level raw score has been determined a simple table lookup with the in-level norms provides the in-level percentile corresponding to this converted in-level raw score. These steps are diagrammed in Figure 7b.



a. In-Level Score Conversion Procedures

b. Out-of-Level Score Conversion Procedures

Figure 7. Score conversion procedures for publishers which provide raw score-to-percentile conversion tables.

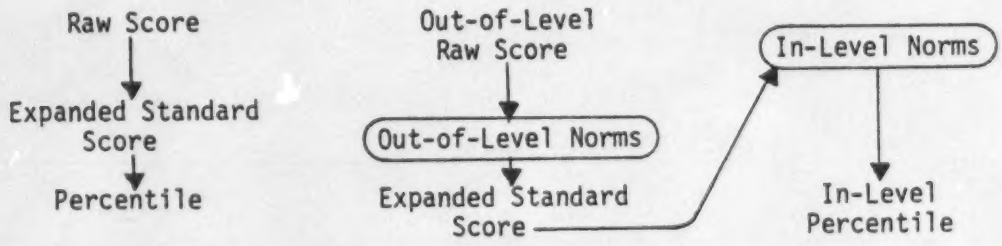
The California Achievement Tests (CAT), Comprehensive Test of Basic Skills (CTBS) and Stanford Achievement Tests (SAT) all provide raw score-to-percentile conversion tables. An example of the steps necessary to complete these conversions is provided in Figure 8 utilizing the Stanford Achievement Tests. In this example the pretest out-of-level raw score of 83 converts to an in-level raw score of 44.5 through the expanded standard score (called Scale Score) of 123. A sample worksheet for converting out-of-level raw scores with this variation for the SAT is given in Appendix C. A blank worksheet is also provided and can be used for any test which has raw score-to-percentile conversion tables.

Variation II. This variation is used when publishers provide expanded standard score-to-percentile conversion tables. In this variation when testing in-level, the raw score must first be converted to an expanded standard score before it can be converted to a percentile (Figure 9a). However, when out-of-level testing is conducted the same three steps are used with two different sets of norms tables. The out-of-level raw score is converted to an expanded standard score using the out-of-level (actual administration level) norms. Once the expanded standard score has been determined,

	Pretest	Posttest
DESCRIPTION		
Test Name:	Stanford Achievement Test	
Subtest Name:	Total Reading	
Expanded Standard Score Name:	Scale Score	
Student Placement:	Grade 4.2	Grade 4.7
Out-of-level instrument administered:	Primary II Level (appropriate for grades 2.7 and 3.2)	Primary II Level (appropriate for grades 2.7 and 3.2)
In-level instrument recommended by test publisher:	Primary III Level which is appropriate for grades 3.7 and 4.2.	Intermediate I Level, which is appropriate for grades 4.7 and 5.2.
PROCEDURES		
1.	Convert the out-of-level raw pretest scores to scale scores using Primary II Level tables. (As an example, use the out-of-level raw score of 83 which converts to the scale score of 123).	1. Convert the out-of-level raw posttest scores to scale scores using Primary II Level tables. (As an example, use the out-of-level raw score of 83 which converts to the scale score of 123).
2.	Using the Primary III Level tables for Beginning of Grade 4 which are recommended by the test publisher for grade 4.2, convert each pretest scale score to its corresponding in-level raw score. (e.g., 44.5)	2. Using the Intermediate I Level tables for End of Grade 4 which are recommended by the test publisher for grade 4.7, convert each posttest scale score to its corresponding in-level raw score. (e.g., 31.5)
3.	Using the Primary III Level raw-score-to-percentile table for Beginning of Grade 4 convert each pretest in-level raw score to its corresponding percentile: (e.g., 11.5)	3. Using the Intermediate I Level raw-scale-to-percentile table for End of Grade 4, convert each posttest in-level raw score to its corresponding percentile. (e.g., 5)

Figure 8. Example of the score conversion procedures for publishers which provide raw score-to-percentile conversion tables.

it is entered directly into the in-level (nominally recommended level) norms to find the corresponding in-level percentile. Theoretically, the in-level percentile is the percentile equivalent which would have been obtained had the in-level test been administered and no floor or ceiling effects encountered. These steps are diagrammed in Figure 9b.



a. In-Level Score Conversion Procedures

b. Out-of-Level Score Conversion Procedures

Figure 9. Score conversion procedures for publishers which provide expanded standard score-to-percentile conversion tables.

The Iowa Tests of Basic Skills (ITBS), Metropolitan Achievement Tests (MAT), Sequential Tests of Educational Progress (STEP II), and SRA Assessment Series all provide expanded standard score-to-percentile conversion tables. An example of the steps necessary to complete these conversions are provided in Figure 10 utilizing the Metropolitan Achievement Tests as an example. A sample worksheet for converting out-of-level raw scores with this variation for the MAT is given in Appendix D. A blank worksheet is also provided and can be used for any tests which have expanded standard score-to-percentile conversion tables.

Checking Publisher Scores. When the scoring services of publishers are utilized it is probably wise to spot check the returned results to be sure that the proper in-level conversions have been made. This can be done using the conversion procedures explained above for Variations I and II.

Pretest	Posttest
DESCRIPTION	
Test Name: Metropolitan Achievement Tests	
Subtest Name: Total Reading	
Expanded Standard Score Name: Standard Score	
Student Placement: Grade 4.2	Grade 4.7
Out-of-level instrument administered: Primary II Level (appropriate for grades 2.7 and 3.1)	Primary II Level (appropriate for grades 2.7 and 3.1)
In-level instrument recommended by test publisher: Elementary Level, which is appropriate for grades 3.7, 4.1 & 4.7.	Elementary Level, which is appropriate for grades 3.7, 4.1, and 4.7.
PROCEDURES	
1. Convert the out-of-level raw pretest scores to standard scores using Primary II Level tables. (As an example, use the out-of-level raw score of 55 which converts to the standard score of 52).	1. Convert the out-of-level raw posttest scores to standard scores using Primary II Level tables. (As an example, use the out-of-level raw score of 73 which converts to the standard score of 59).
2. Using the Elementary Level standard score-to-percentile table for Beginning of Grade 4 convert each pretest in-level standard score to its corresponding percentile. (e.g., 16)	2. Using the Elementary Level standard score-to-percentile table for End of Grade 4, convert each posttest in-level standard score to its corresponding percentile. (e.g., 20)

Figure 10. Example of the score conversion procedures for publishers which provide expanded standard score-to-percentile conversion tables.

References

Ayrer, J.E. and McNamara, T.C. Survey testing on an out-of-level basis. Journal of Educational Measurement, 1973, 10, 79-84.

CTB/McGraw Hill. California Achievement Test: Test Coordinators' Handbook Preliminary Edition, Forms C and D, Levels 10-19. Monterey, California: CTB/McGraw Hill, 1977.

Davis, F.B. Item selection techniques. In E.F. Lindquist (Ed.), Educational Measurement. Washington, D.C.: American Council on Education, 1951.

Fisher, J.A. The use of out-of-grade tests with retarded and accelerated readers. Doctoral dissertation, University of Iowa, 1961. (University Microfilms No. 61-5564).

Gulliksen, H. Theory of mental tests. New York: Wiley, 1950.

Hieronymous, A.N. and Lindquist, E.F. Manual for administrators, supervisors, and counselors, Levels Edition, Forms 5 and 6, Iowa Tests of Basic Skills. Boston: Houghton Mifflin Co., 1974.

Keating, D.P. Testing those in the top percentiles. Exceptional Children, 1975, 41, 435-436.

Long, J.V., Schaffran, J.A., and Kellogg, T.M. Effects of out-of-level survey testing on reading achievement scores of Title I, ESEA students. Journal of Educational Measurement, 1977, 14, 203-213.

Prescott, G.A. Metropolitan Achievement Tests: Manual for Interpreting. New York: Harcourt Brace Jovanovich, 1972.

Roberts, A.H. Out of Level Testing. Technical Paper No. 6 of the Title I Evaluation and Reporting System. Mountainview, California: RMC Research Corp., October 1976.

Stanley, J.C. and Hopkins, K.D. Educational and psychological measurement and evaluation. Englewood Cliffs, N.J.: Prentice-Hall, Inc., 1972.

Teachers College Press. Extended (out-of-level) norms for the Gates-MacGinite Reading Tests. New York: Teachers College Press, 1977.

Yoshida, R.K. Out-of-level testing of special education students with a standardized achievement battery. Journal of Educational Measurement, 1976, 13, 215-221.

Young, L.R. An investigation of the scaling of the Metropolitan Achievement Tests. Doctoral dissertation. Fordham University, 1972. (University Microfilms No. 73-1521).

Reference Notes

1. Wick, J.W. Testing students at functioning reading level: A two-year report from Chicago. Manuscript submitted for publication, 1977.

2. Wick, J.W. A report on individualized testing in Chicago. In Putting Current Ideas in Measurement to Work in the Schools. Proceedings from the Thirteenth Annual Houghton Mifflin Measurement Conference, Iowa City, Iowa, April 26-27, 1977.

3. Mendro, R. A report on individualized testing in Dallas. In Putting Current Ideas in Measurement to Work in the Schools. Proceedings from the Thirteenth Annual Houghton Mifflin Measurement Conference, Iowa City, Iowa, April 26-27, 1977.

4. Barnes, J. A report on individualized testing in Atlanta. In Putting Current Ideas in Measurement to Work in the Schools. Proceedings from the Thirteenth Annual Houghton Mifflin Measurement Conference, Iowa City, Iowa, April 26-27, 1977.

5. Hieronymus, A.N. Individualized testing with the Iowa Tests of Basic Skills. In Putting Current Ideas in Measurement to Work in the Schools. Proceedings from the Thirteenth Annual Houghton Mifflin Measurement Conference, Iowa City, Iowa, April 26-27, 1977.

6. Pelavin, S.H. and Barker, P. The generalizability of the results of a standardized achievement test. Paper presented at the Annual Meeting of the American Educational Research Association, 1976.

7. Tallmadge, G.K. and Roberts, A.O.H. Factors that influence test results: ESEA Title I Evaluation and Reporting System. Mountainview, California: RMC Research Corporation. Manuscript submitted to U.S. Office of Education, January 1978.

8. This formula implies that responses are purely at random. In many instances one or more of the alternatives can be eliminated even when the respondent does not know the correct answer. A number of formulae have been presented to handle this problem of guessing. See Davis (1951) for a discussion of this topic.

9. Roberts, S.J. Test floor and ceiling effects: ESEA Title I Evaluation and Reporting System. Mountainview, Calif.: RMC Research Corp. Manuscript submitted to U.S. Office of Education, January 1978.

10. Note that the multilevel editions of the Iowa Tests of Basic Skills and the SRA Achievement Series produced some extremely high percentiles at the lowest tabled level. A portion of these discrepancies can be accounted for by the wide range of in-level tabled norms which are available and the unlikelihood that a test level considerably above the student's nominal level would be administered to a student who is functioning considerably below his/her nominal level.

11. Note that the authors are not encouraging overtesting, but more appropriate testing. Any testing which can maximally benefit the individual student is not a waste of that student's time. Therefore, general achievement testing can often be used most beneficially to evaluate program outcomes and direct individual student analysis.

APPENDIX A

Chance Score Grade Equivalents and Percentiles of Various Standardized Tests

This appendix contains the chance raw scores and their corresponding grade equivalent scores and percentile ranks for reading and mathematics subtests of tests which are commonly used in program evaluation for grades 2-6. Indicated for the subtests of each level of each test appropriate for this grade span are the total number of items, the number of alternatives per item, the chance raw score and the grade equivalent corresponding to this chance raw score. The percentile equivalents of this chance score for the lowest and highest in-level tabled values available for grades between the beginning of grade 2 (B-2) and the end of grade 6 (E-6) are also provided. Where the number of alternatives varied among items within a subtest, the number of items is indicated for each number of alternatives. In some instances, a level was designed to go below or above the indicated tabled values for percentiles. The nominally recommended administration levels for each test and level are given on the following page. These nominally recommended levels should especially be considered when examining tests which have multilevel editions, such as the Iowa Tests of Basic Skills and the SRA Achievement Series.

Nominally Recommended Grade Levels for Test Levels of Ten Achievement Tests Widely Used in Grades 2-6

Test Name	Level Name	Nominally Recommended Grades Lowest Grade	Highest Grade
California Achievement Tests (1977)	12	1.5*	2.9
	13	2.5	3.9
	14	3.5	4.9
	15	4.5	5.9
	16	5.5	6.9
Comprehensive Tests of Basic Skills (1973)	C	1.6	2.9
	1	2.5	4.9
	2	4.5	6.9
Gates-MacGinitie Reading Tests (1971)	Primary B	B-2**	E-2***
	Primary C	B-3	E-3
	Survey D	B-4	E-6
Iowa Tests of Basic Skills (1972)	7	1.7	2.5
	8	2.6	3.5
	9	B-3	E-3
	10	B-4	E-4
	11	B-5	E-5
	12	B-6	E-6
Metropolitan Achievement Tests (1970)	Primary I	1.5	2.4
	Primary II	2.5	3.4
	Elementary	3.5	4.9
	Intermediate	5.0	6.9
SRA Achievement Series	Primary I	B-1	B-2
	Primary II	E-2	B-4
	Multilevel (Blue)	B-4	E-6
	Multilevel (Green)	B-6	B-8
Stanford Achievement Test (1973)	Primary I	1.5	2.4
	Primary II	2.5	3.4
	Primary III	3.5	4.4
	Intermediate I	4.5	5.4
	Intermediate II	5.5	6.9
Stanford Diagnostic Reading Test (1976)	Red	1.6	3.5
	Green	2.6	5.5
	Brown	4.6	9.5
Stanford Diagnostic Mathematics Test (1976)	Red	1.6	4.5
	Green	3.6	6.5
Sequential Tests of Educational Progress II (1969)	4	E-3	B-6

* First digit indicates grade; second digit indicates month within grade where September is month 0.
** B = beginning of grade.
*** E = end of grade.

California Achievement Tests, Level 12 (Form C, 1977)	Number of Items	Number of Alternatives	Chance Raw Score	Grade Equiv.	Grade	%ile	Grade	%ile
Phonic Analysis	25	10/3 15/4	7	-	B-2	10	B-3	4
Structural Analysis	11	4	3	-	B-2	11	B-3	2
Reading Vocabulary	15	4	4	1.4	B-2	18	B-3	7
Reading Comprehension	20	4	5	1.6	B-2	21	B-3	7
Total Reading	71	10/3 61/4	19	1.1	B-2	8	B-3	1
Language Mechanics	20	4	5	1.2	B-2	9	B-3	2
Language Expression	20	2	10	1.5	B-2	19	B-3	7
Total Language	40	20/2 20/4	15	1.1	B-2	9	B-3	2
Math Computation	26	4	7	1.6	B-2	23	B-3	3
Math Concepts and Applications	40	4	10	1.7	B-2	3	B-3	1
Total Mathematics	66	4	17	1.1	B-2	5	B-3	1

California Achievement Tests, Level 13 (Form C, 1977)	Number of Items	Number of Alternatives	Chance Raw Score	Grade Equiv.	Grade	%ile	Grade	%ile
Phonic Analysis	20	5/3 15/4	5	-	E-2	9	B-4	-
Structural Analysis	11	4	3	-	E-2	10	B-4	-
Reading Vocabulary	15	5/3 10/4	4	1.6	E-2	13	B-4	3
Reading Comprehension	27	7/3 20/4	7	1.5	E-2	35	B-4	2
Total Reading	73	17/3 36/4	20	1.5	E-2	8	B-4	2
Math Computation	40	5	8	2.3	E-2	25	B-4	2
Math Concepts and Applications	45	5	9	1.6	E-2	11	B-4	3
Total Mathematics	85	5	17	1.7	E-2	11	B-4	2
Language Mechanics	20	10/4 10/5	5	1.8	E-2	16	B-4	4
Language Expression	26	11/2 11/3 4/4	10	1.7	E-2	14	B-4	3
Total Language	46	11/2 11/3 14/4 10/5	15	1.7	E-2	13	B-4	2

California Achievement Tests, Level 14 (Form C, 1977)	Number of Items	Number of Alternatives	Chance Raw Score	Grade Equiv.	Grade	%ile	Grade	%ile
Reading Vocabulary	30	5/3 25/4	8	2.6	E-3	25	B-5	6
Reading Comprehension	40	4	10	2.5	E-3	22	B-5	7
Total Reading	70	5/3 65/4	18	2.6	E-3	22	B-5	5
Language Mechanics	25	5	5	2.1	E-3	12	B-5	4
Language Expression	38	8/2 12/3 18/4	13	2.3	E-3	18	B-5	7
Total Language	63	8/2 12/3 18/4 25/5	18	2.3	E-3	13	B-5	4
Math Computation	40	5	8	3.0	E-3	15	B-5	2
Math Concepts and Applications	45	5	9	2.4	E-3	14	B-5	3
Total Mathematics	85	5	17	2.5	E-3	11	B-5	2

California Achievement Tests, Level 15 (Form C, 1977)	Number of Items	Number of Alternatives	Chance Raw Score	Grade Equiv.	Grade	%ile	Grade	%ile
Reading Vocabulary	30	5/3 25/4	8	2.9	E-4	12	B-6	5
Reading Comprehension	40	4	10	2.8	E-4	16	B-6	6
Total Reading	70	5/3 65/4	18	2.8	E-4	12	B-6	4
Language Mechanics	25	5	5	2.4	E-4	8	B-6	4
Language Expression	38	8/2 12/3 18/4	13	2.3	E-4	9	B-6	4
Total Language	63	8/2 12/3 18/4 25/5	18	2.4	E-4	6	B-6	3
Math Computation	40	5	8	3.5	E-4	11	B-6	2
Math Concepts and Applications	45	5	9	3.0	E-4	12	B-6	4
Total Mathematics	85	5	17	3.2	E-4	8	B-6	2

California Achievement Tests, Level 16 (Form C, 1977)	Number of Items	Number of Alternatives	Chance Raw Score	Grade Equiv.	Grade	%ile	Grade	%ile
Reading Vocabulary	30	5/3 25/4	8	3.4	E-5	11	E-6	6
Reading Comprehension	40	4	10	3.4	E-5	13	E-6	8
Total Reading	70	5/3 65/4	18	3.3	E-5	10	E-6	5
Language Mechanics	25	5	5	2.3	E-5	4	E-6	4
Language Expression	38	8/2 12/3 18/4	13	2.5	E-5	7	E-6	6
Total Language	63	8/2 12/3 18/4 25/5	18	2.5	E-5	4	E-6	3
Math Computation	40	5	8	3.9	E-5	9	E-6	4
Math Concepts and Applications	45	5	9	3.2	E-5	8	E-6	4
Total Mathematics	85	5	17	3.5	E-5	6	E-6	3

Comprehensive Test of Basic Skills, Level C (Form S, 1973)	Number of Items	Number of Alternatives	Chance Raw Score	Grade Equiv.	Grade	%ile	Grade	%ile
Vocabulary	33	4	8	--	B-2	6	E-2	2
Sentences	23	4	6	--	B-2	17	E-2	8
Passages	18	4	5	--	B-2	24	E-2	9
Total Reading	74	4	19	1.3	B-2	6	E-2	3
Expression	22	4	6	--	B-2	15	E-2	7
Spelling	34	2	17	--	B-2	19	E-2	9
Mechanics	23	4	6	--	B-2	23	E-2	12
Total Language	79	34/2 45/4	28	0.7	B-2	13	E-2	4
Math Computation	28	4	7	--	B-2	14	E-2	5
Math Concepts/Applic.	25	4	6	--	B-2	5	E-2	2
Total Mathematics	53	4	13	0.9	B-2	6	E-2	2

Comprehensive Test of Basic Skills, Level 1 (Form S, 1973)	Number of Items	Number of Alternatives	Chance Raw Score	Grade Equiv.	Grade	%ile	Grade	%ile
Vocabulary	40	4	10	1.9	B-2	42	E-4	4
Comprehension	45	4	11	2.1	B-2	50	E-4	6
Total Reading	85	4	21	1.9	B-2	39	E-4	4
Spelling	44	2	22	2.1	B-2	49	E-4	8
Mechanics	20	4	5	1.7	B-2	35	E-4	5
Expression	30	4	8	1.9	B-2	39	E-4	5
Total Language	94	44/2 50/4	35	2.0	B-2	44	E-4	4
Math Computation	48	4	12	2.2	B-2	57	E-4	2
Math Concepts	25	4	6	1.6	B-2	32	E-4	3
Math Applications	25	4	6	2.2	B-2	53	E-4	6
Total Mathematics	98	4	25	1.9	B-2	39	E-4	2

Comprehensive Test of Basic Skills, Level 2 (Form 2, 1973)	Number of Items	Number of Alternatives	Chance Raw Score	Grade Equiv.	Grade	%ile	Grade	%ile
Vocabulary	40	4	10	2.5	B-4	16	E-6	4
Comprehension	45	4	11	2.6	B-4	16	E-6	6
Total Reading	85	4	21	2.3	B-4	12	E-6	3
Spelling	50	2	25	2.6	B-4	19	E-6	8
Mechanics	20	4	5	2.0	B-4	11	E-6	5
Expression	35	4	9	2.1	B-4	10	E-6	5
Total Language	105	50/2 55/4	39	2.3	B-4	12	E-6	3
Math Computation	48	4	12	3.0	B-4	15	E-6	3
Math Concepts	25	4	6	2.0	B-4	9	E-6	3
Math Applications	25	4	6	2.9	B-4	22	E-6	6
Total Mathematics	98	4	25	2.5	B-4	9	E-6	2

Gates MacGinitie Reading Test, Primary B (Forms 1M and 2M, 1971)	Number of Items	Number of Alternatives	Chance Raw Score	Grade Equiv.	Grade	%ile	Grade	%ile
Vocabulary	48	4	12	1.3	B-2	16	E-2	4
Comprehension	34	4	9	1.5	B-2	27	E-2	4

Gates MacGinite Reading Test, Primary C (Forms 1M and 2M, 1971)	Number of Items	Number of Alternatives	Chance Raw Score	Grade Equiv.	Grade	%ile	Grade	%ile
Vocabularly	52	4	13	1.5	B-3	5	E-3	2⁻
Comprehension	48	4	12	1.8	B-3	14	E-3	4

Gates MacGinite Reading Test, Survey D (Forms 1M, 2M and 3M; 1971)	Number of Items	Number of Alternatives	Chance Raw Score	Grade Equiv.	Grade	%ile	Grade	%ile
Vocabulary	50	4	13	3.1	B-4	18	E-6	2⁻
Comprehension	52	4	13	2.9	B-4	18	E-6	3

Iowa Test of Basic Skills, Level 7 (Form 5, 1971)	Number of Items	Number of Alternatives	Chance Raw Score	Grade Equiv.	Grade	%ile	Grade	%ile
Vocabulary	30	4	8	1.0	B-2	11	E-3	1⁻
Word Analysis	49	3	16	0.7	B-2	3	E-3	1⁻
Reading Comprehension	66	32/2 23/3 11/4	26	1.4	B-2	18	E-3	1
Language								
Spelling	27	3	9	1.1	B-2	8	E-3	1
Capitalization	40	*	—	—	—	—	—	—
Punctuation	42	*	—	—	—	—	—	—
Usage	23	3	8	1.1	B-2	14	E-3	1
Mathematics								
Math Concepts	33	3	11	0.8	B-2	5	E-3	1⁻
Math Problems	22	3	7	0.8	B-2	3	E-3	1⁻

Iowa Test of Basic Skills, Level 8 (Form 5, 1971)	Number of Items	Number of Alternatives	Chance Raw Score	Grade Equiv.	Grade	%ile	Grade	%ile
Vocabulary	30	13/3 17/4	9	1.6	B-2	32	E-3	3
Word Analysis	58	3	19	1.2	B-2	18	E-3	1
Reading Comprehension	67	16/2 28/3 23/4	23	1.7	B-2	33	E-3	2
Language								
Spelling	29	3	10	1.8	B-2	37	E-3	5
Capitalization		*	—	—	—	—	—	—
Punctuation		*	—	—	—	—	—	—
Usage	23	3	8	1.7	B-2	38	E-3	5
Mathematics								
Math Concepts	36	3/3 33/4	9	1.4	B-2	12	E-3	1⁻
Math Problems	30	5/3 25/4	8	1.2	B-2	18	E-3	1⁻

* Number of responses varies by item.

Iowa Test of Basic Skills, Level 9 (Form 5, 1971)	Number of Items	Number of Alternatives	Chance Raw Score	Grade Equiv.	Grade	%ile	Grade	%ile
Language								
Spelling	31	5	6	2.2	B-3	20	E-6	1⁻
Capitalization	38	4	10	2.5	B-3	31	E-6	1⁻
Punctuation	38	4	10	2.9	B-3	42	E-6	2
Usage	32	4	8	2.2	B-3	26	E-6	1⁻
Mathematics								
Math Concepts	30	4	8	2.3	B-3	15	E-6	1⁻
Math Prob. Solving	25	4	6	2.2	B-3	21	E-6	1⁻

Iowa Test of Basic Skills, Level 10 (Form 5, 1971)	Number of Items	Number of Alternatives	Chance Raw Score	Grade Equiv.	Grade	%ile	Grade	%ile
Language								
Spelling	38	5	8	2.9	B-3	44	E-6	2
Capitalization	39	4	10	2.9	B-3	44	E-6	1
Punctuation	39	4	10	3.2	B-3	52	E-6	4
Usage	32	4	8	2.7	B-3	39	E-6	1⁻
Mathematics								
Math Concepts	36	4	9	3.2	B-3	53	E-6	1⁻
Math Prob. Solving	27	4	7	3.2	B-3	53	E-6	1⁻

Iowa Test of Basic Skills, Level 11 (Form 5, 1971)	Number of Items	Number of Alternatives	Chance Raw Score	Grade Equiv.	Grade	%ile	Grade	%ile
Language								
Spelling	43	5	9	3.4	B-3	59	E-6	1⁻
Capitalization	40	4	10	3.4	B-3	58	E-6	1⁻
Punctuation	40	4	10	3.5	B-3	61	E-6	1⁻
Usage	32	4	8	3.3	B-3	54	E-6	1⁻
Mathematics								
Math Concepts	42	4	11	3.8	B-3	74	E-6	1
Math Prob. Solving	29	4	7	3.7	B-3	77	E-6	3

Iowa Test of Basic Skills, Level 12 (Form 5, 1971)	Number of Items	Number of Alternatives	Chance Raw Score	Grade Equiv.	Grade	%ile	Grade	%ile
Language								
Spelling	46	5	9	4.0	B-3	72	E-6	1⁻
Capitalization	42	4	11	4.5	B-3	83	E-6	1⁻
Punctuation	42	4	11	4.4	B-3	80	E-6	1⁻
Usage	32	4	8	3.8	B-3	64	E-6	1⁻
Mathematics								
Math Concepts	45	4	11	4.6	B-3	94	E-6	1⁻
Math Prob. Solving	31	4	8	5.0	B-3	97	E-6	1⁻

Metropolitan Achievement Test, Primary I (Form F, 1970)	Number of Items	Number of Alternatives	Chance Raw Score	Grade Equiv.	Grade	%ile	Grade	%ile
Word Knowledge	35	4	9	1.1	B-2	1		
Word Analysis	40	4	10	1.1	B-2	1		
Reading	42	3	14	1.4	B-2	11		
Total Math	62	33/0* 29/5	6	1.0⁻	B-2	1		

Metropolitan Achievement Test, Primary II (Form F, 1970)	Number of Items	Number of Alternatives	Chance Raw Score	Grade Equiv.	Grade	%ile	Grade	%ile
Word Knowledge	40	4	10	1.7	B-2	24	E-3	2
Word Analysis	35	4	9	1.3	B-2	8	E-3	1
Reading	44	3	15	1.7	B-2	30	E-3	5
Math Computation	33	5	7	1.3	B-2	20	E-3	1⁻
Math Concepts	40	5	8	1.1	B-2	6	E-3	1⁻
Math Prob. Solving	35	5	7	1.3	B-2	20	E-3	1⁻

Metropolitan Achievement Test, Elementary (Form F, 1970)	Number of Items	Number of Alternatives	Chance Raw Score	Grade Equiv.	Grade	%ile	Grade	%ile
Word Knowledge	50	4	13	2.2	B-3	14	E-4	2
Reading	45	4	11	1.9	B-3	10	E-4	2
Math Computation	40	5	8	1.9	B-3	7	E-4	1
Math Concepts	40	5	8	2.0	B-3	11	E-4	1
Math Prob. Solving	35	5	7	2.1	B-3	10	E-4	1
Language	50	15/4 35/5	11	2.3	B-3	26	E-4	3

Metropolitan Achievement Test, Intermediate (Form F, 1970)	Number of Items	Number of Alternatives	Chance Raw Score	Grade Equiv.	Grade	%ile	Grade	%ile
Word Knowledge	50	4	13	3.2	B-5	11	E-6	3
Reading	45	4	11	3.1	B-5	14	E-6	2
Math Computation	40	5	8	3.8	B-5	8	E-6	1
Math Concepts	40	5	8	3.3	B-5	8	E-6	2
Math Prob. Solving	35	5	7	3.2	B-5	10	E-6	2
Language	103	38/4 65/5	23	3.2	B-5	12	E-6	2

*Item format was open-ended; no alternatives were presented to examine.

SRA Achievement Series, Primary I (Form F, 1971)	Number of Items	Number of Alternatives	Chance Raw Score	Grade Equiv.	Grade	%ile	Grade	%ile
Total Language Arts	63	19/2 17/3 27/4	22	1.0	B-2	3	E-3	1
Total Reading	43	10/3 33/4	12	1.0	B-2	2	E-3	1
Total Mathematics	53	4	13	1.0	B-2	1	E-3	1

SRA Achievement Series, Primary II (Form F, 1971)	Number of Items	Number of Alternatives	Chance Raw Score	Grade Equiv.	Grade	%ile	Grade	%ile
Total Language Arts	65	15/2 16/3 34/4	21	1.3	B-2	14	E-4	1
Total Reading	52	4	13	1.6	B-2	29	E-4	2
Total Mathematics	58	4	13	1.6	B-2	20	E-4	1

SRA Achievement Series, Multilevel Edition (Blue, Form E 1971)	Number of Items	Number of Alternatives	Chance Raw Score	Grade Equiv.	Grade	%ile	Grade	%ile
Reading Comprehension	60	4	15	3.4	E-3	39	B-9	2
Vocabulary	30	4	8	2.1	E-3	9	B-9	1
Total Reading	90	4	23	2.6	E-3	21	B-9	1
Math Concepts	40	4	10	2.9	E-3	22	B-9	1
Math Computation	40	5	8	3.1	E-3	21	B-9	1
Total Mathematics	80	40/4 40/5	18	3.1	E-3	18	B-9	1
Grammatical usage	70	14/2 13/3 33/4 10/5	22	2.7	E-3	24	B-9	1
Spelling	40	5	8	2.6	E-3	13	B-9	1
Total Language Arts	110	14/2 13/3 33/4 30/5	30	2.6	E-3	17	B-9	1

SRA Achievement Series, Multilevel Edition (Green, Form E, 1971)	Number of Items	Number of Alternatives	Chance Raw Score	Grade Equiv.	Grade	%ile	Grade	%ile
Reading Comprehension	48	4	12	3.6	E-4	27	E-9	2
Vocabulary	42	4	11	3.4	E-4	24	E-9	1
Total Reading	90	4	23	3.4	E-4	24	E-9	1
Grammatical Usage	70	14/2 12/3 34/4 10/5	22	3.1	E-4	17	E-9	1
Spelling	40	5	8	3.8	E-4	30	E-9	2
Total Language Arts	110	14/2 12/3 34/4 50/5	30	3.4	E-4	20	E-9	1
Math Concepts	40	4	10	3.6	E-4	21	E-9	2
Math Computation	40	5	8	3.7	E-4	19	E-9	2
Total Mathematics	80	40/4 40/5	18	3.7	E-4	17	E-9	1

Stanford Achievement Test, Primary Level I (Form A, 1973)	Number of Items	Number of Alternatives	Chance Raw Score	Grade Equiv.	Grade	%ile	Grade	%ile
Vocabulary	37	3	12	K.1	B-2	6		
Reading (Part A)	45	3	15	1.0	B-2	8		
Reading (Part B)	42	3	14	1.2	B-2	11		
Word Study Skills	60	3	20	K.9	B-2	2		
Total Reading	147	3	49	K.8	B-2	4		
Math Concepts	32	4	8	K.1	B-2	1		
Math Computation	32	4	8	K.6	B-2	4		
Total Mathematics	64	4	16	K.8	B-2	1		

Stanford Achievement Test, Primary Level II (Form A, 1973)	Number of Items	Number of Alternatives	Chance Raw Score	Grade Equiv.	Grade	%ile	Grade	%ile
Vocabulary	37	3	12	1.0	E-2	4	B-3	2
Reading (Part A)	45	3	15	1.4	E-2	4	B-3	4
Reading (Part B)	48	4	12	1.6	E-2	6	B-3	6
Word Study Skills	65	3	22	1.3	E-2	4	B-3	1
Total Reading	158	110/3 48/4	49	1.2	E-2	2	B-3	2
Math Concepts	35	1/3 31/4 1/5 2/23	8	1.0	E-2	2	B-3	1
Math Computation	37	8/3 29/5	8	1.0	E-2	2	B-3	2
Math Applications	28	11/4 17/5	6	1.3	E-2	4	B-3	2
Total Mathematics	100	9/3 42/4 47/5 2/23	23	1.0	E-2	1	B-3	1

Stanford Achievement Test, Primary Level III (Form A, 1973)	Number of Items	Number of Alternatives	Chance Raw Score	Grade Equiv.	Grade	%ile	Grade	%ile
Vocabulary	45	4	11	1.0	E-3	4	B-4	2
Reading Comprehension	70	4	18	1.7	E-3	4	B-4	4
Word Study Skills	55	31/3 24/4	16	1.8	E-3	6	B-4	10
Total Reading	125	31/3 94/4	34	1.6	E-3	4	B-4	4
Math Concepts	32	4	8	2.1	E-3	8	B-4	8
Math Computation	36	24/3 12/5	10	2.2	E-3	8	B-4	8
Math Applications	28	5	6	2.1	E-3	8	B-4	8
Total Mathematics	96	24/3 32/4 40/5	24	2.0	E-3	4	B-4	4

Stanford Achievement Test, Intermediate Level I (Form A, 1973)	Number of Items	Number of Alternatives	Chance Raw Score	Grade Equiv.	Grade	%ile	Grade	%ile
Vocabularly	50	4	13	2.2	E-4	6	B-5	4
Reading Comprehension	72	4	18	2.7	E-4	11	B-5	10
Word Study Skills	55	25/3 30/4	16	2.1	E-4	10	B-5	10
Total Reading	127	25/3 102/4	34	2.5	E-4	8	B-5	6
Math Concepts	32	4	8	3.0	E-4	12	B-5	10
Math Computation	40	18/3 22/5	10	3.2	E-4	10	B-5	10
Math Applications	40	5	8	2.1	E-4	6	B-5	4
Total Mathematics	112	18/3 32/4 62/5	26	2.5	E-4	4	B-5	4

Stanford Achievement Test, Intermediate Level II (Form A, 1973)	Number of Items	Number of Alternatives	Chance Raw Score	Grade Equiv.	Grade	%ile	Grade	%ile
Vocabularly	50	4	13	3.4	E-5	8	E-6	4
Reading Comprehension	71	4	18	3.3	E-5	11	E-6	6
Word Study Skills	50	25/4 25/5	11	3.0⁻	E-5	10	E-6	8
Total Reading	121	96/4 25/5	29	3.0⁻	E-5	8	E-6	6
Math Concepts	35	4	9	3.5	E-5	6	E-6	4
Math Computation	45	24/3 21/5	12	3.9	E-5	10	E-6	6
Math Applications	40	5	8	3.0⁻	E-5	8	E-6	4
Total Mathematics	120	24/3 35/4 61/5	29	3.2	E-5	4	E-6	2

Stanford Diagnostic Mathematics Test, Red Level (Form A, 1976)	Number of Items	Number of Alternatives	Chance Raw Score	Grade Equiv.	Grade	%ile	Grade	%ile
Number System and Numeration	30	5	6	K	2	6	4	1⁻
Computation	33	5	7	1.2	2	7	4	1⁻
Applications	30	5	6	1.0	2	3	4	1⁻
Mathematics Total	93	5	19	1.4	2	3	4	1⁻

Stanford Diagnostic Mathematics Test, Green Level (Form A, 1976)	Number of Items	Number of Alternatives	Chance Raw Score	Grade Equiv.	Grade	%ile	Grade	%ile
Number System and Numeration	36	5	7	2.6	4	6	6	1
Computation	48	5	10	2.5	4	11	6	1
Applications	30	5	6	2.6	4	12	6	3
Mathematics Total	114	5	23	2.5	4	7	6	1

Stanford Diagnostic Reading Test, Red Level (Form A, 1976)	Number of Items	Number of Alternatives	Chance Raw Score	Grade Equiv.	Grade	%ile	Grade	%ile
Auditory Vocabularly	36	3	12	1.3	2	3⁻	3	1⁻
Auditory Discrimination	40	4	10	K	2	1⁻	3	1⁻
Phonetic Analysis	40	4	10	K	2	2	3	1
Word Reading	42	3	14	K	2	11	3	3
Reading Comprehension	48	3	16	K	2	12	3	4
Comprehension Total	90	3	30	1.4	2	11	3	3

Stanford Diagnostic Reading Test, Green Level (Form A, 1976)	Number of Items	Number of Alternatives	Chance Raw Score	Grade Equiv.	Grade	%ile	Grade	%ile
Auditory Vocabularly	40	3	13	2.6	3	6	5	2
Auditory Discrimination	36	3	12	1.3	3	7	5	3
Phonetic Analysis	36	3	12	2.0	3	14	5	8
Structural Analysis	60	3	20	K	3	5	5	2
Literal Comprehension	30	4	8	K	3	8	5	2
Inferential Comprehension	30	24/3 6/4	10	K	3	12	5	3
Reading Comprehension Total	60	24/3 36/4	17	2.5	3	7	5	2

Stanford Diagnostic Reading Test, Brown Level (Form A, 1976)	Number of Items	Number of Alternatives	Chance Raw Score	Grade Equiv.	Grade	%ile	Grade	%ile
Auditory Vocabularly	40	3	13	3.3	5	9	6	6
Literal Comprehension	30	4	8	2.9	5	18	6	11
Inferential Comprehension	30	4	8	2.6	5	20	6	12
Reading Comprehension Total	60	4	15	2.9	5	14	6	8
Phonetic Analysis	36	3	12	2.4	5	13	6	11
Structural Analysis	34	24/V* 30/4	13	K	5	1	6	1-
Reading Rate	34	3	11**	4.2**	5	42**	6	29**

* Number of possible choices varied with each item in this section.
** Rapidly timed subtest which tests the reading rate (number of items completed) of each student. A chance score is not meaningful for such tests.

APPENDIX B

NAMES OF EXPANDED STANDARD SCORE SCALES
USED BY MAJOR TEST PUBLISHERS

Test	Name of Expanded Standard Score Scale
California Achievement Tests (1970)	Achievement Development Scale Scores
California Achievement Tests (1977)	Scale Scores
Comprehensive Tests of Basic Skills (1968, 1973)	Expanded Standard Scores
Gates-MacGinitie Reading Tests (1971)	Does Not Presently Provide Expanded Standard Scores*
Iowa Tests of Basic Skills	Grade Equivalents
Iowa Test of Educational Development	Growth Scale
Metropolitan Achievement Tests	Standard Scores
Sequential Tests of Educational Progress II	Converted Scores
SRA Assessment Survey	Growth Scale
Stanford Achievement Test	Scale Scores

*Provides special norms for tests administered one year below grade level. These norms can be used at grades 2, 3, and 4 using the Primary Edition and at grades 7 and 10 using the Survey Edition.

APPENDIX C

WORKSHEET FOR CONVERTING OUT-OF-LEVEL TEST SCORES

Used When Publishers Provide Raw Score-to-Percentile Conversion Tables

Test: Stanford Achievement Test, Form A
Subtest: Total Reading
Grade: 8
Date Administered: October 15, 1977

Out-of-level Test Administered: Intermediate II
In-Level Test Recommended by Publisher: Advanced
Expanded Standard Score Scale Name: Scaled Scores

Student Name	Out-of-Level Raw Score	Expanded Standard Score from Out-of-Level Norms	Raw Score from In-Level Norms	Percentile Rank from In-Level Norms
Source of Score	Number Right	Raw Score-to-Scaled Score conversion table, Intermediate II Manual, Part II, pp. 29-30	Raw Score-to-Scaled Score conversion table, Advanced Manual, Part II, pp. 28-29	Scaled Score-to-Percentile Rank conversion table, Advanced Manual, Part II, p. 25 (be sure to use percentile rank tables for grade and time of year test administered)
	EXAMPLE			
Sample 1	83	175	60	70
Sample 2	54	155	41	16
Sample 3	65	162	47.5 (midpoint of interval 47-48)	24.3 (obtained by interpolation)

WORKSHEET FOR CONVERTING OUT-OF-LEVEL TEST SCORES

Used When Publishers Provide Raw Score-to-Percentile Conversion Tables

Test:
Subtest:
Grade:
Date Administered:

Out-of-Level Test Administered:
In-Level Test Recommended by Publisher:
Expanded Standard Score Scale Name:

Student Name	Out-of-Level Raw Score	Expanded Standard Score from Out-of-Level Norms	Raw Score from In-Level Norms	Percentile Rank from In-Level Norms

APPENDIX D

WORKSHEET FOR CONVERTING OUT-OF-LEVEL TEST SCORES

Used When Publishers Provide Expanded Standard Score-to-Percentile Conversion Tables

Test: Metropolitan Achievement Test
Subtest: Total Reading
Grade: 5
Date Administered: October 15, 1977

Out-of-Level Test Administered: Elementary, Form F
In-Level Test Recommended by Publisher: Intermediate, Form F
Expanded Standard Score Scale Name: Standard Scores

Student Name	Out-of-Level Raw Score	Expanded Standard Score from Out-of-Level Norms	Percentile Rank from In-Level Norms
Source of Score	Number Right	Raw Score-to-Standard Score conversion table, Teacher's Handbook, Elementary Level, p. 6	Standard Score-to-Percentile Rank conversion table, Teacher's Handbook, Intermediate Level, p. 8 (be sure to use percentile rank tables for grade and time of year test was administered)
EXAMPLE			
Sample 1	45	58	12
Sample 2	60	64	24
Sample 3	67	69	34

WORKSHEET FOR CONVERTING OUT-OF-LEVEL TEST SCORES

Used When Publishers Provide Expanded Standard Score-to-Percentile Conversion Tables

Test:
Subtest:
Grade:

Out-of-Level Test Administered:
In-Level Test Recommended by Publisher:
Expanded Standard Score Scale Name:

Student Name	Out-of-Level Raw Score	Expanded Standard Score from Out-of-Level Norms	Percentile Rank from In-Level Norms